Exploring Disease Protein Aggregation with Coarse Grained Models

Qualifying Exam Proposal (Draft)
Nicolas Fawzi
October 29, 2004
UCSF/UC Berkeley Joint Graduate Group in Bioengineering

Subjects for Examination

Protein structure, function, chemistry
Molecular simulations
Aggregation disease biology
Statistical Mechanics

Hypothesis

The hydrophobic sequence patterns that influence protein topology and direct folding are critical determinants of aggregation intermediate structures, pathways and kinetics of proteins L and G and disease proteins amyloid β-peptide and transthyretin.

Specific Aims

The medical relevance of protein aggregation is clear – it plays a significant role in poorly understood diseases (such as Alzheimer's, Huntington's and Mad Cow), and also presents one of the most significant hurdles in the efficient production of stable protein biopharmaceuticals. In protein production, genetically engineered proteins expressed in bacteria often accumulate in inclusion bodies, insoluble amorphous aggregates of inactive and misfolded proteins. Though the formation of inclusion bodies and macroscopic amyloid fibrils, large ordered aggregates associated with diseases, can be investigated through current experimental techniques, early events in protein aggregation involving the formation of pre-fibrillar species (small aggregates of a few protein chains that are thought to be capable of nucleating the large, characteristic amyloid fibrils) are too small, too quick and too disordered for experimental probing by most current bio-chemical techniques or protein structural techniques (x-ray crystallography and NMR). As pointed out in a recent review, "understanding of the pathway and mechanism of amyloidogenesis is important for the development of useful therapeutic strategies for these disease" [1]. The research I propose to undertake will precisely address this currently unsolved problem.

The primary goal of my work will be to develop a physics-based model of protein aggregation that reproduces known experimental aggregation characteristics and enables predictions of aggregation rates and pathways for disease proteins and re-engineered mutants. In order to achieve this goal, I will continue development of a sequence based, protein folding model created in the Head-Gordon lab. We have already demonstrated that this model is capable of reproducing the key aspects in protein folding. In a study comparing two proteins with the same topology but experimentally known different folding mechanisms, the model replicated distinct folding mechanisms based only on differences in the amino acid sequence. Since protein aggregation is a property generic to the amino acid chain and believed to known to depend on the protein sequence, this model provides us with the ability to recreate the major physical aspects of protein aggregation while simulating the long timescales over which aggregation events occur. I will improve this model by adding an explicit representation of hydrogen bonding, a key interaction in the formation of β-sheet structure, known to be the major structural component of

amyloid fibrils. Other groups have proposed coarse-grained models to study aggregation, however, their models rely on lattice models, Go potentials which encode protein native state information, or physics based models lacking the ability to represent native topologies. As of today, none of these models is sufficient for describing the details of both folding and aggregation behavior, a necessity for understand the transition from folded to aggregated states that is involved in disease.

In order to examine disease protein aggregation, we will first apply our model to less complex system protein systems and pursue direct experimental validation. We have already completed computational studies on protein L and G with our current coarse-grained model where we conclude that designing proteins which have increased diffuse contacts in the unfolded state and fast-forming intermediates abates aggregation. To validate these models, we propose two strategies for this validation – the first by simulating different simple peptides, and the second by completing a joint experimental/computational study on an extensive mutant library of proteins L and G that we have begun. Simulations of simple peptide systems will confirm that the model is capable of reproducing aggregation behavior and aide the model parameterization. Joint efforts on proteins L and G will confirm the model's ability to recreate aggregation events for entire protein systems, and enable us to determine critical sequence and structural factors that govern protein aggregation.

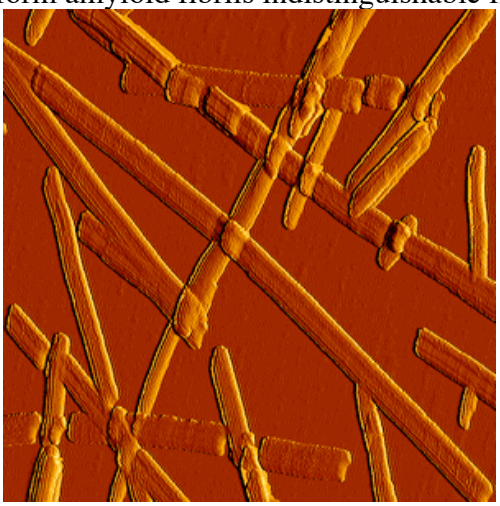
With this validated model, I will study the aggregation of two proteins involved in aggregation diseases – amyloid β (Aβ) and transthyretin (TTR), the main aggregating species in Alzheimer's disease and various familial associated pathologies, respectively. Each of these proteins in known to aggregate *in vivo* and *in vivo*, however, the mechanism of aggregation and specifically the steps along the path to aggregation, have not yet been determined. Working with a computationally efficient model, I will simulate over the entire aggregation process of Aβ and TTR, from short formation of short amyloid fibrils, and create an all-atom model of the fibrillar structure. Comparing the structure of the simulated aggregates to known amyloid characteristics will demonstrate that the model is capable of replicating true aggregation events. With this connection established, the model can then be used to elucidate the steps between unaggregated and aggregated, where experimental techniques cannot yet probe. By working with this theoretical model, I will investigate the effects of sequence mutations and solution conditions on aggregation, propose pathways and structures of early aggregation on the molecular level, and compare and verify experimental findings of aggregation mechanism.

Background

Protein Folding, Misfolding and Amyloid Fibrils

Nature has designed proteins to fold to a specific three-dimensional structure and function as a structural and/or chemical building block in living systems. The sequence of amino acids encodes this native structure as well as the "energy landscape" on which the protein searches out conformations. The functional native state of the protein most often corresponds to the thermodynamic (free-energy) minimum conformation at physiological conditions [2]. The discovery of protein aggregation diseases, however, where multiple proteins sacrifice contacts in the globular native state in favor of inter-chain contacts with neighboring proteins, suggests that some proteins have aggregated states that are thermodynamically equally if not more favorable than the native state. Many of these aggregates have a common morphology named amyloid fibrils [3] – regular fibrillar structures micrometers in length, a few nanometers in diameter composed of an intramolecular β-sheet core running perpendicular to the fibril axis. It has more

recently been discovered that under a variety of solution conditions, most proteins can be made to form amyloid fibrils indistinguishable from that of disease proteins [4].


a

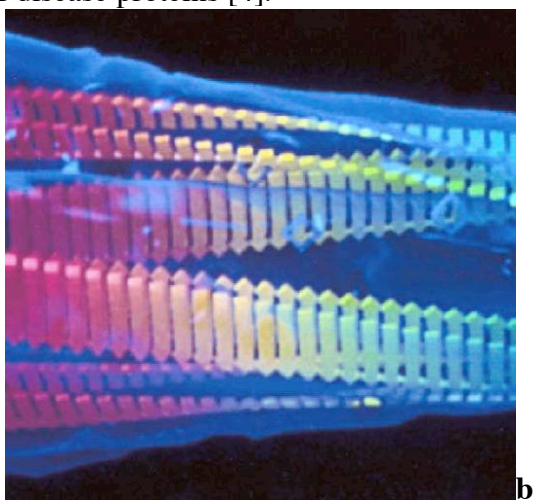

b

Figure 1 Amyloid Fibrils
a AFM image of Amyloid Fibrils (JPK Instruments AG)
b Schematic of cross β-structure in amyloid fibrils [5]

There are at least 16 distinct human diseases that are associated with amyloid fibril formation [6]. Significant advancements in research in the past five years have led to the discovery that the toxic species in the amyloid diseases may not be the fibrils themselves, but rather the pre-fibrillar aggregates [7]. Although early attention focused on the possible toxicity of the amyloid fibrils, it is now hypothesized that the early aggregates are the main toxic species in aggregates [8], underscoring the need to develop an understanding of the entire aggregation process, not simply the structure of the final amyloid fibril.

Alzheimer's Disease

Alzheimer's disease is a neurodegenerative disease linked to the aggregation and amyloid fibril formation of a short peptide, β amyloid (also referred to as Aβ), created by proteolytic cleavage of the Aβ precursor protein (APP) [9]. This fragment contains part of the C-terminal region of the protein, and this peptide is known to be highly prone to aggregate *in vitro* and *in vivo* to form amyloid fibrils associated with Alzheimer's Disease (AD). The structure of this peptide in the monomeric state is unstable and has at least two meta-stable states under various conditions – α-helical and polyproline(II) [10, 11]. The monomers then aggregate to form the amyloid fibrils from any of these states, showing that a Go model description of aggregation is not sufficient. The mechanism of aggregation of Aβ monomers is thought to proceed through the formation of a high free-energy thermodynamic nucleus of a handful of chains, after which the aggregation proceeds energetically downhill. The structure of this nucleus, however, has not been determined. The final Aβ amyloid fibrils have been extensively studied by Tycko and coworkers who have published a model structure based on constraints from NMR [12].

Transthyretin-associated aggregation diseases

Transthyretin (TTR) is a 127 amino acid homotetramer known to be involved in the number of aggregation diseases including senile systemic amyloidosis and familial amyloidotic

neuropathy. Both of these diseases involve the formation and accumulation of amyloid fibrils of TTR. Many TTR associated diseases have been tied to at least 80 point mutations in the TTR sequence, leading to increased propensity of contracting a TTR associated amyloid diseases in certain segments of the population. Aggregation of TTR and its disease mutants is slow under physiological conditions, and the rate-limiting step for aggregation is the dissociation of the tetramer. After tetramer dissociation, the monomers must then sacrifice native interactions for contacts in the aggregated state.

Recent work from Jeffrey W. Kelly's group suggests that the mechanism of aggregation for TTR monomers might be a complex downhill reaction, distinct from the nucleation dependent reaction currently accepted for Aβ [13]. Serag and coworkers also recently proposed a strand arrangement for one region of the TTR fibrils from site directed spin labeling techniques [14], shedding some of the first light on the molecular structure of an amyloid fibril from a disease protein. As with Aβ aggregation, the pathway to aggregation of TTR and the structural characteristic of the protein aggregates along the way have yet to be determined.

Models of protein folding and aggregation

The Head-Gordon lab has developed a model of protein folding and aggregation that can be applied to answer open questions about aggregation mechanism of disease proteins. The original model, described in detail in previous papers [15-18], is an off-lattice α carbon (C_α) bead model of a protein inspired by the work of Thirumalai and co-workers [19-21]. The model is general to proteins with any α-helical, β-sheet, or mixed α and β topologies. The model represents the complex set of molecular interactions critical to protein folding and aggregation (hydrophobic, electrostatic, van der Waals, hydrogen bonds, solvent effects) with a minimal potential energy function [22]. Each amino acid in the protein chain is represented by a bead with a hydrophobic (B), hydrophilic (L), or neutral (N) interaction character. In order to simulate interactions favoring the collapse of non-polar amino acids into a hydrophobic core, the interaction between hydrophobic amino acids is represented by a standard Lennard-Jones potential. In the current model, regions of the sequence are biased towards, but not assigned, a particular secondary structure through a dihedral angle potential energy function for four adjacent beads.

Although other groups have developed minimalist model studies of aggregation, they appear to make assumptions that are incompatible with certain experimentally known characteristics of aggregation phenomena for disease proteins. Due to their high computational efficiency, lattice models have been used to study protein folding aggregation by a number of groups. These models, however, lack the ability to faithfully describe the complex secondary and tertiary structures, as well as being unsuited to directly interrogate folding and aggregating kinetics. Other groups have focused on Go models. In their studies of SH3 and Aβ aggregation, Shakhnovich and Peng and coworkers rely on Go potentials to stabilize a native state and aggregated state [23, 24]. Traditionally, Go potentials specify the native state for the protein by giving favorable interactions for contacts present in the native state, and unfavorable interactions for any non-native contacts. In these studies, the authors expanded Go potentials to aggregation and inter-protein interactions by giving favorable interactions to residues on different proteins, "assuming that if two amino acid residues that attract to each other in a single protein [native state] will also have attraction in different proteins" [23]. Since a Go model has favorable energy only at near-native distances, these models rely on a very strong (3× the Go interaction

strength) generic hydrogen bonding term to create non-specific interchain interactions. This accounts for the type of aggregates that are created in these simulations – nearly completely extended chains, with most of the protein involved in β-sheet pairing. True amyloid fibrils, however, show a diversity in the percentage of the amino acids involved in the repeating β-sheet structure [3]. This spectrum is likely caused by differences in sequence between amyloid forming proteins.

Our model, by comparison, maintains a physics-based potential between amino acids. Because of this, we were able to replicate the different folding mechanisms for proteins L and G with the same immunoglobulin-binding topology by changing only the amino acid sequence. A Go model is not be able to distinguish between proteins of the same topology and will not be able to capture both the folded and the amyloid state of proteins, the two stable states of proteins in aggregation events. Our study will recreate formation of fibrillar species from Aβ and TTR monomers without relying on Go type native-state potentials.

Methods

Completed Goal – *Develop and characterize a coarse-grained model of protein folding*

The coarse-grained model developed in the Head-Gordon lab over the past few years has been extensively characterized and described in several publications. Much of the work on the model has focused on proteins L and G, members of the immunoglobulin fold class and adopt the same α/β fold topology but with experimentally confirmed different folding pathways [25]. In a previous paper that I contributed to, we demonstrated that these models of protein L and G reproduce the distinct folding pathways of these two proteins based on their distinct sequences. Additionally, we show that this model is general to different folds by successfully modeling the WW Domain fold by mapping the amino acid sequence to our coarse grained sequence and performing a few rounds of design (as in L and G) to account for ambiguity in the mapping. This success suggests that this model captures essential interactions in protein folding. Since the interactions that stabilize aggregates are the same as those involved in folding, we are confident the model will be able to recreate protein aggregation.

Completed Goal – *Demonstrate the ability to model protein aggregation. Learn general properties that relate folding characteristics to aggregation.*

With this first generation model, we recently completed a study linking the folding properties of a protein to its aggregation behavior. Protein L and G, though they have the same topology, have different folding kinetics and stationary points (transition, intermediate and denatured state ensembles). In this work and previous work, we simulate and thoroughly characterize these stationary points as well as the kinetics and thermodynamics of aggregation. By examining contacts formed in these single-chain states, we can draw conclusions on the formation of intra and interchain contacts formed in the aggregated states. Figure 2 presents an example of the contact maps for protein L and protein G that support our conclusions. Though protein G folds slower than protein L, protein G aggregates slower than L due to its rapidly formed folding intermediate which exhibits mostly native contacts spread across the protein chain. Though the formation of the native state proceeds slowly from this intermediate, the quick protection of hydrophobic regions of the protein results in slower aggregation.

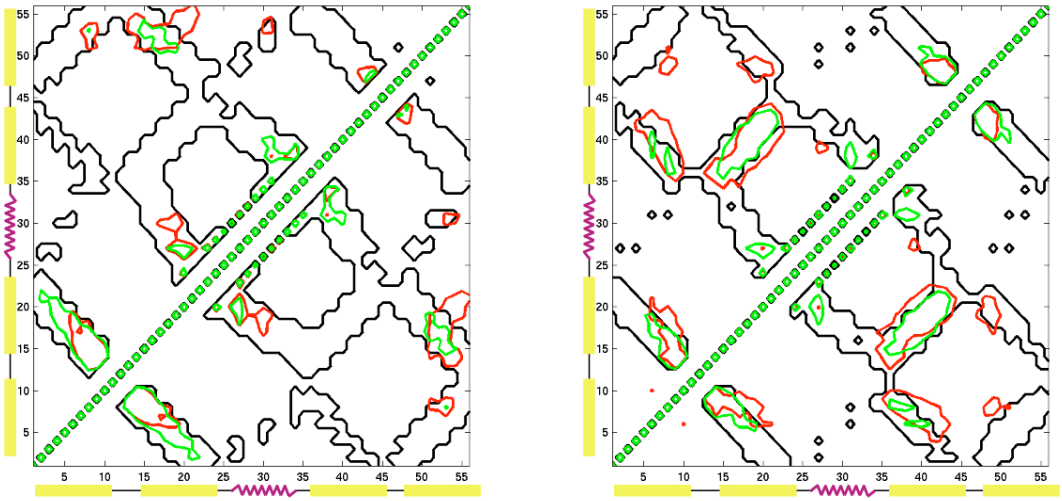


Figure 2 Intrachain contact maps in protein L and G aggregation, comparing native state (black) and contacts that are present across at least 60% of the denatured state for L or intermediate ensemble for G (red), and intra-chain contacts made in at least 60% of the aggregated ensemble (green) for (a) protein L's slow aggregation pathway, (b) protein G's slow aggregation pathway [26]. Secondary structural elements of the native sequence are displayed on the left and bottom of the maps. The overlap of the intrachain contacts with the contacts in the denatured/intermediate state indicates that the protein aggregates from this state. The greater spread of contacts in G's intermediate state compared to L's denatured state account for the slower G aggregation rate.

This work demonstrated our ability to connect specific folding properties to aggregation rates. At the same time, we realize that the model is still limited in the type of aggregation it can represent. Protein aggregates found in diseases are often amyloid fibrils with a very specific cross β-sheet geometry. Most coarse-grained models, including our first generation model, are not able to faithfully represent β-sheets due to difficulty representing the hydrogen bonding that stabilizes the sheet geometry. In order to study disease protein aggregation, I will incorporate this hydrogen-bond functionality into the model, validate this new model, and use the model for predictions.

Goal – *Complete an improved minimalist model for aggregation including a generalized hydrogen bonding term*

A coarse-grained model is designed to reproduce key interactions necessary to represent the folding and aggregation of a protein chain. Previous minimalist models have relied upon specific native state interactions, Go potentials or local dihedral propensities to form secondary structures. In order to improve the model, I will complete the addition of a hydrogen bonding term to replicate this interaction that is critical for secondary structure and amyloid fibrils formation. Inspired by the Mercedes Benz model of water [27], I have already added a hydrogen bonding term that forms β-sheet structure, general to any combination of parallel and anti-parallel strands. The effect of this potential has been to form idealized β sheet structures. This functionality of the model was tested in the most recent Critical Assessment of Techniques for Protein Structure Prediction (CASP) to generate initial configurations for protein structure prediction optimization. Figure 3a is an example of an all-atom structure of a protein that was

created from a starting structure generated using the coarse-grained model. I have also completed some preliminary tests of the new hydrogen bonding function under aggregation conditions for Protein L and G. Figure 3b is an example of the aggregates created using this potential.

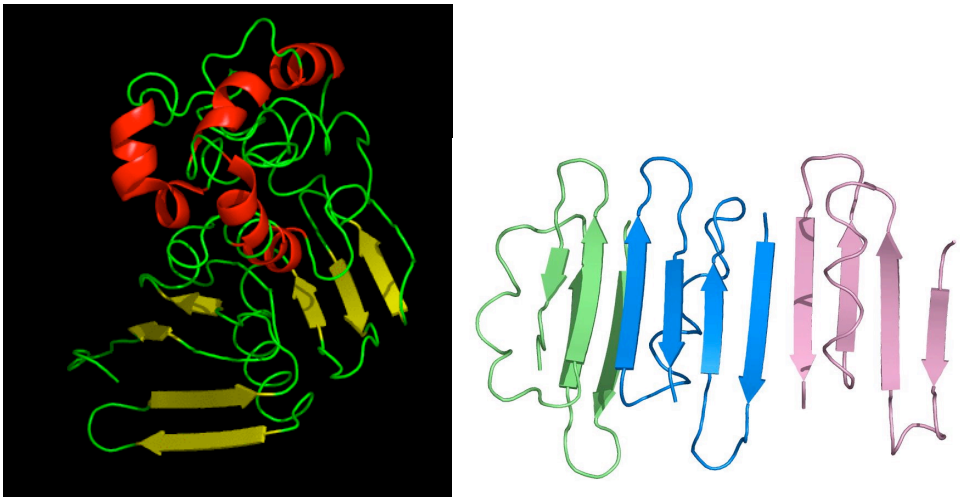


Figure 3 Hydrogen Bonding Potential – Preliminary tests
a CASP6 prediction created from a coarse-grained model initial structure
b Fibril like β-sheet assembly from the coarse-grained model

The current minimalist model I am working with gives certain regions of the chain a propensity to form a certain type of secondary structure. In order to faithfully replicate protein aggregation in general, we will seek to develop a model where areas of the chain will be able to form coil, sheet and helix conformations based on the local bead-sequence. This ability will be useful for faithfully replicating the aggregation of TTR, where the aggregated state is thought to have less β-sheet content than the native state, most likely indicating a β-sheet to coil conversion [14]. Looking to future studies, ability to change secondary structure would be critical for modeling prion protein aggregation where it is believed that a portion of the protein in α-helical conformation converts to from a β-sheet with another chain, forming the basis of an aggregate [28].

I am already testing the new potential in conjunction with the current minimalist model potentials and have preliminary results that indicate it maintains or improves the cooperativity of folding based on established metrics (heat capacity peak width, folding temperature range based on % folded) enables a more faithful mapping of the 20 letter amino acid code to the three letter minimalist bead sequence. I plan to extend this "hydrogen bond" potential to add a minimum favorable for α-helical arrangements. The effect of these potentials combined will be to mimic hydrogen bond formation in either α-helix or β-sheet secondary structure arrangement.

Comparing the Aggregation Model to Experiment

Goal – *Compare aggregation behavior of peptide sequences in vitro and in silico*

In order to be confident in the performance of the model, we can compare the aggregation characteristics of simulated and actual peptides. In a recent review, Zhang has compiled a list of related peptides that have different aggregation behavior [29]. Most of these peptides are based

on hydrophobic-hydrophilic amino acid patterning, known to be critical to β-sheet formation. I have identified one set of three distinct but related peptides that will be of particular interest to examine (Table 1).

Peptide name	Sequence	Aggregation Behavior
EFK8-II	n-FEFKFEFK-c	β-sheet aggregates
EAKA8-I	n-AEAKAEAK-c	no aggregation
EAKA16-I	n-AEAKAEAKAEAKAEAK-c	β-sheet aggregates

Table 1 Peptides with known aggregation behavior [29]

The EFK8-II peptide readily forms aggregates with β-sheet structure, while a peptide with the four phenylalanine residues replaced with alanine, EAKA8-I, displays no aggregation at the same conditions. Simulating the aggregation of these two peptides in our model could be a check to see that the model can distinguish between sequences with different aggregation behavior. The distinction between these two peptides should certainly be accessible to the model since replacing F with A significantly lowers the hydrophobicity of the peptide, an interaction that we model explicitly. In order to confirm that any simulated difference in aggregation is significant, we can also compare EAK8A-I with EAKA16-I, where the 8 residues of the EAKA8-I peptide are repeated. If the model can reproduce the difference in aggregation behavior based solely on the length of the peptide and the conformational and entropic effects the likely cause this difference, we can proceed with the knowledge that the model is effectively simulating the underlying forces driving aggregation.

Goal – *Simulate the aggregation of protein L and G, predict the aggregation of protein L and G mutants, confirm with experiment*

To truly test our ability to model protein aggregation, we will go beyond the simple peptide tests proposed above by making predictions concerning the aggregation of protein L and G and related mutants. In conjunction with the Blanch lab, we will conduct complementary experiments and simulations on these proteins are proposed to examine the effect of mutating protein sequence on nucleation events, aggregation propensity, the kinetics of aggregation and folding, and the role of folding intermediates on aggregation. Proteins L and G provide an excellent experimental and computational system to examine these effects on aggregation due to their similar topology but different folding mechanism, and their extensive experimental characterization. By comparing the folding and aggregation of a library of computational mutants to corresponding library of experimental mutants, we hope to validate the coarse-grained model while at the same time elucidating the sequence and structural factors that governs aggregation events in greater detail than possible in experiment alone. Once validated by experiment, simulations will provide a rapid screening for sequences that minimize aggregation, and will aid in the design of protein sequences that exhibit aggregation inhibition. Using our results and those from other protein engineering studies on protein L, we will construct a set of guidelines for the rational design of mutations for reducing the aggregation propensity of any protein. To test the transferability of these guidelines, we will study the aggregation propensity of a wide-range of mutants of proteins G.

Simulation of Aβ Aggregation

Goal – *Create a model of amyloid fibril aggregation for Aβ, a peptide of medical interest.*

By simulating the aggregation of multiple chains, we anticipate seeing intramolecular β-sheet formation that mimics Aβ fibrils. In order to compare these results to experiments, we can calculate a diffraction pattern using an elastic diffraction formula [24]. In addition to generating a fibril like assembly with the minimalist model and calculating a diffraction pattern from this structure, we can also convert the amyloid fibril to an all atom structure and perform local or near-local optimization. We know that the minimalist model can recreate most of the essential elements of protein folding and yet does not perfectly preserve the interactions and the sizes of an all atom model; therefore we can convert to all atom space in order to generate a more faithful diffraction pattern. This serves a few purposes – to give our best guess at the structure of the fibrils, and to convince those who do not believe that minimalist models can recreate the most important aspects of protein aggregation. If we can generate a plausible all atom model of an amyloid fibril, it lends credence to our proposed mechanism of amyloid formation and the conclusions we draw about aggregation kinetics and thermodynamic properties from the minimalist model dynamics. With large enough systems (~30 peptides) we can determine if the short fibrils formed agree with experiment.

Goal – *Confirm experimentally known thermodynamic and kinetic aggregation properties. Propose a thermodynamic nucleus and aggregation pathway.*

Aggregation rates are dependent on a number of factors, including both the amino acid sequences as well as the solution conditions. Although we cannot get an absolute prediction of aggregation rate from our minimalist models, we can predict relative aggregation rates for mutant sequences and solution conditions including temperature, monomer concentration and simulated denaturant. We can simulate the aggregation under various conditions and mutations and check the degree to which the relative rates determined by experiment match our findings.

Simulation techniques are now starting to be used to study these aggregation phenomena, requiring that the research we pursue to add to what has already been done. In addition to using a physically relevant model, the computational efficiency of the model enables us to fully characterize the aggregating systems. We plan to use the multiple histogram method to search for the thermodynamic nucleus of aggregation, one of the major challenges in understanding aggregation that is not yet possible experimentally. In order to understand how aggregation proceeds, we would like to know what structures give rise to the amyloid fibrils we see after aggregation occurs [1]. These transition state structures are likely to be short-lived, small, (thought to be on the order of three to ten chains) and retain key areas of structural similarity without a rigid "native" structure. Techniques we have developed enable us to characterize these structures in our simulations. We will extend the method of searching for transition state structures already developed to be used for aggregation studies [30]. If we can have an idea of the number of chains involved and their arrangement, we can propose to experimentally search for these species and check to see if these are the path critical structures between unaggregated monomers and fibrils.

Simulation of TTR aggregation

Goal – *Characterize aggregation structures, kinetics and thermodynamics of TTR protein and compare to Aβ aggregation. Propose general properties of aggregation and amyloid formation events.*

At this point, little is known about the mechanism of fibrillization, particularly if different mechanisms can lead to the familiar amyloid structure. There are currently some suggestions that studying one model of amyloid formation is not enough. Jeffrey Kelly and his group have recently published findings claiming that the aggregation of TTR is a downhill process, distinct from the thermodynamic nucleation mechanism currently supported for Aβ aggregation. Additionally, Serag and coworkers have recently suggested a strand ordering in the TTR fibril that resembles portions of the native state [14, 31]. Taken together, these findings suggest that though both Aβ and TTR form amyloid fibrils, the mechanism for reaching this fibril is distinct, and therefore the crucial steps involved in aggregation may be completely different. If we seek to understand what events are critical for protein aggregation in order to change aggregation propensity or rate, we will need to understand the aggregation mechanism for that particular protein or one with a similar fibrillization process, not for any generic protein.

With the model, I will simulate the aggregation of TTR and compare it to Aβ. I will attempt to confirm that TTR aggregation is indeed a downhill process and to demonstrate the formation of short fibrils consistent with the current facts related to strand arrangement. With these connection points to experiment, I will investigate the conversion from native TTR to aggregated TTR and search for the transition state in these conversions. Having a model for the transition state between folded and aggregated, we can propose a site for a pharmaceutical approach to changing TTR aggregation through a small molecule stabilizing the folded state or disrupts critical contacts in the transition state, hence slowing aggregation [32].

Conclusions

With a faithful model of the interactions present in proteins, we can model aggregation and gain insight into the sequences and structures that are critical for disease protein aggregation. Along the way, we will develop and validate the model with experimental comparison, and test the model's predictive power. Having established a connection to experimental knowledge of disease protein aggregation, we will propose structures and pathways in protein aggregation that are inaccessible to current experimental techniques.

References

1. Thirumalai, D., D.K. Klimov, and R.I. Dima, *Emerging ideas on the molecular basis of protein and peptide aggregation*. Current Opinion in Structural Biology, 2003. **13**(2): p. 146-159.
2. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels*. Nature Structural Biology, 1997. **4**(1): p. 10-19.
3. Dobson, C.M., *Protein folding and misfolding*. Nature, 2003. **426**(6968): p. 884-890.
4. Rochet, J.C. and P.T. Lansbury, *Amyloid fibrillogenesis: themes and variations*. Current Opinion in Structural Biology, 2000. **10**(1): p. 60-68.
5. Dobson, C.M., *Principles of protein folding, misfolding and aggregation*. Seminars in Cell & Developmental Biology, 2004. **15**(1): p. 3-16.
6. Kelly, J.W., *The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways*. Current Opinion in Structural Biology, 1998. **8**(1): p. 101-106.

7. Stefani, M. and C.M. Dobson, *Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution*. Journal of Molecular Medicine-Jmm, 2003. **81**(11): p. 678-699.
8. Bucciantini, M., et al., *Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases*. Nature, 2002. **416**(6880): p. 507-511.
9. Cohen, F.E. and J.W. Kelly, *Therapeutic approaches to protein-misfolding diseases*. Nature, 2003. **426**(6968): p. 905-909.
10. Eker, F., K. Griebenow, and R. Schweitzer-Stenner, *A beta(1-28) fragment of the amyloid peptide predominantly adopts a polyproline II conformation in an acidic solution*. Biochemistry, 2004. **43**(22): p. 6893-6898.
11. Tiana, G., et al., *Thermodynamics of beta-amyloid fibril formation*. Journal of Chemical Physics, 2004. **120**(17): p. 8307-8317.
12. Petkova, A.T., et al., *A structural model for Alzheimer's beta-amyloid fibrils based on experimental constraints from solid state NMR*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(26): p. 16742-16747.
13. Hurshman, A.R., et al., *Transthyretin aggregation under partially denaturing conditions is a downhill polymerization*. Biochemistry, 2004. **43**(23): p. 7365-7381.
14. Serag, A.A., et al., *Arrangement of subunits and ordering of beta-strands in an amyloid sheet*. Nature Structural Biology, 2002. **9**(10): p. 734-739.
15. Sorenson, J.M. and T. Head-Gordon, *Protein engineering study of protein L by simulation*. Journal of Computational Biology, 2002. **9**(1): p. 35-54.
16. Sorenson, J.M. and T. Head-Gordon, *The importance of hydration for the kinetics and thermodynamics of protein folding: simplified lattice models*. Folding & Design, 1998. **3**(6): p. 523-534.
17. Sorensen, J.M. and T. Head-Gordon, *Toward minimalist models of larger proteins: A ubiquitin-like protein*. Proteins-Structure Function and Genetics, 2002. **46**(4): p. 368-379.
18. Sorenson, J.M. and T. Head-Gordon, *Redesigning the hydrophobic core of a model beta-sheet protein: Destabilizing traps through a threading approach*. Proteins-Structure Function and Genetics, 1999. **37**(4): p. 582-591.
19. Honeycutt, J.D. and D. Thirumalai, *The Nature of Folded States of Globular-Proteins*. Biopolymers, 1992. **32**(6): p. 695-709.
20. Klimov, D.K. and D. Thirumalai, *Factors governing the foldability of proteins*. Proteins-Structure Function and Genetics, 1996. **26**(4): p. 411-441.
21. Guo, Z. and D. Thirumalai, *Kinetics and thermodynamics of folding of a de Novo designed four-helix bundle protein*. Journal of Molecular Biology, 1996. **263**(2): p. 323-343.
22. Dill, K.A., *Dominant Forces in Protein Folding*. Biochemistry, 1990. **29**(31): p. 7133-7155.
23. Ding, F., et al., *Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism*. Journal of Molecular Biology, 2002. **324**(4): p. 851-857.
24. Peng, S., et al., *Discrete molecular dynamics simulations of peptide aggregation*. Physical Review E, 2004. **69**(4).
25. Brown, S., N.J. Fawzi, and T. Head-Gordon, *Coarse-grained sequences for protein folding and design*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(19): p. 10712-10717.

26. Fawzi, N.J., et al., *Influence of Denatured and Intermediate States of Folding on Protein Aggregation*. submitted, 2004.
27. Silverstein, K.A.T., A.D.J. Haymet, and K.A. Dill, *A simple model of water and the hydrophobic effect*. Journal of the American Chemical Society, 1998. **120**(13): p. 3166-3175.
28. Pan, K.M., et al., *Conversion of Alpha-Helices into Beta-Sheets Features in the Formation of the Scrapie Prion Proteins*. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(23): p. 10962-10966.
29. Zhang, S.G., *Emerging biological materials through molecular self-assembly*. Biotechnology Advances, 2002. **20**(5-6): p. 321-339.
30. Brown, S. and T. Head-Gordon, *Intermediates and the folding of proteins L and G*. Protein Science, 2004. **13**(4): p. 958-970.
31. Serag, A.A., et al., *Identification of a subunit interface in transthyretin amyloid fibrils: Evidence for self-assembly from oligomeric building blocks*. Biochemistry, 2001. **40**(31): p. 9089-9096.
32. Hammarstrom, P., et al., *Prevention of transthyretin amyloid disease by changing protein misfolding energetics*. Science, 2003. **299**(5607): p. 713-716.